Available at
www.ElsevierMathematics.com
POWERED BY SCIENCE DIRECT®
Journal of Multivariate Analysis 91 (2004) 240–261
Journal of Multivariate Analysis
http://www.elsevier.com/locate/jmva

Bandwidth choice for local polynomial estimation of smooth boundaries

Peter Hall[a] and Byeong U. Park[a,b,*,1]

[a] *Centre for Mathematics and its Applications, Australian National University, Canberra, ACT 0200, Australia*
[b] *Department of Statistics, Seoul National University, Seoul 151–742, South Korea*

Received 3 December 2001

Abstract

Local polynomial methods hold considerable promise for boundary estimation, where they offer unmatched flexibility and adaptivity. Most rival techniques provide only a single order of approximation; local polynomial approaches allow any order desired. Their more conventional rivals, for example high-order kernel methods in the context of regression, do not have attractive versions in the case of boundary estimation. However, the adoption of local polynomial methods for boundary estimation is inhibited by lack of knowledge about their properties, in particular about the manner in which they are influenced by bandwidth; and by the absence of techniques for empirical bandwidth choice. In the present paper we detail the way in which bandwidth selection determines mean squared error of local polynomial boundary estimators, showing that it is substantially more complex than in regression settings. For example, asymptotic formulae for bias and variance contributions to mean squared error no longer decompose into monotone functions of bandwidth. Nevertheless, once these properties are understood, relatively simple empirical bandwidth selection methods can be developed. We suggest a new approach to both local and global bandwidth choice, and describe its properties.

AMS 2000 subject classifications: primary 62G05; secondary 62G09

Keywords: Bandwidth choice; Bootstrap; Global bandwidth; Local bandwidth; Local polynomial methods; Nonparametric curve estimation

*Corresponding author. Fax: 822-883-6144.
E-mail address: bupark@stats.snu.ac.kr (B.U. Park).
[1] Supported in part by Korea Research Foundation Grant (KRF-2000-015-DP0059), and by the Brain Korea 21 Project. Computational assistance of Jun Hyeok Hwang is greatly appreciated.

0047-259X/$ - see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2003.10.002

1. Introduction

Local polynomial methods for nonparametric curve estimation have been very influential, not least on account of their extraordinary flexibility and adaptivity. Their advantages in the context of regression and related problems are legion; see for example [7]. There is a variety of ways of applying them to density estimation, for example by converting the density estimation problem to one of regression [19] or by using local likelihood techniques [5,13,18,20]. As well having a multitude of applications to exclusively nonparametric problems, local polynomial methods have been used in a variety of parametric settings where nonparametric methods are used to provide enhancement; see for example [1,2,4,14].

Local polynomial techniques have potentially a great deal to offer in problems of boundary estimation too, where they promise a particularly flexible approach in a context where existing solutions are usually very restricted. In particular, estimators based on data envelope analysis or DEA [8] are necessarily of second order, and so have the same convergence rates as local linear methods [9,15]. Free disposal hull or FDH estimators [6] are similarly restricted; see [3,16,17,21]. In contrast, local polynomial methods have the potential to supply to boundary estimation the unsurpassed degree of flexibility and adaptivity they offer in other settings.

Local polynomial boundary estimators were introduced by Hall et al. [11] as a development of parametric techniques of the same general type. The structure of their limiting distributions is unknown, however, and neither are formulae for their optimal bandwidths. Nothing is known about practical methods for empirical bandwidth choice. The present paper resolves these issues, and focuses particularly on the bandwidth choice problem. Matters such as limiting distribution are raised in order to shed light on properties of bandwidth.

Local constant estimators have been discussed in an interesting paper by Gijbels and Peng [10]. The construction of their estimator differs from ours in important aspects, and in particular involves two smoothing parameters (their h_n and m) rather than our single h. The choices of Gijbels and Peng's (2000) h_n and m that are permitted by their theory do not allow their estimator to enjoy quite as good a convergence rate as our local constant estimator, discussed in Section 2, but it seems possible to overcome this problem by reframing the conditions and conducting the proof a little differently. The context of Gijbels and Peng (2000) is that of a nonrandom number of independent and identically distributed data, rather than a Poisson number of data, but this does not affect either the definition of the estimator or its convergence rate. See Section 2 for further discussion of these issues.

In more conventional problems of bandwidth choice the mean squared error of an estimator decomposes, to first order, into a part (representing squared bias) that is a strictly increasing function of bandwidth, and another (derived from variance) that is strictly decreasing in the bandwidth. Indeed, insofar as their dependence on bandwidth is concerned the bias and variance components are proportional to h^a and h^{-b}, respectively, where h denotes the bandwidth and $a, b > 0$. In such cases the formula for the asymptotically optimal bandwidth is simple to derive, and easy to

use as the basis for empirical calculations. The context of local polynomial estimation of smooth boundaries is far more complex, however. Formulae for mean squared error no longer decompose into monotone functions of bandwidth. Moreover, local polynomial estimators in this setting are highly nonlinear functions of the data, and so the matter of convergence of moments, essential to a study of mean squared error, is relatively complex. Furthermore, the limiting distributions are more closely related to the exponential than they are to the normal, although even there the links are more apparent in terms of context (e.g. connections to Poisson processes) than through mathematical formulae (expressions for limiting distributions turn out to be very case-dependent in the context of local polynomial methods).

In Sections 2 and 3 we give detailed descriptions of large-sample theory for local constant and local linear estimators, respectively. Results for higher-order cases are similar, differing only in the rapidly increasing complexity of the limiting distribution. In Section 4 we use the insight obtained from these properties to develop a general bootstrap approach to empirical bandwidth choice, and to describe its properties. Our technique is applicable to both local and global bandwidth choice. It involves a new method for bootstrapping inhomogeneous Poisson processes, not involving explicit estimation of the intensity function. Mathematical details behind our arguments are given in Section 5.

2. Local constant boundary estimator

The boundary curve, $\mathscr{C}$, will be assumed to have Cartesian representation $y = g(x)$. Suppose data pairs (X_i, Y_i), comprising a dataset $\mathscr{P}$, are generated by a Poisson process with intensity $n\mu(\cdot,\cdot)$ in the plane $\mathbb{R}^2$. Assume $\mu(x,y) = 0$ for $y > g(x)$, and that μ is bounded away from 0 and is continuous in a neighbourhood, below $\mathscr{C}$, of the boundary point $(x_0, g(x_0))$. These assumptions will be made throughout Sections 2 and 3. We wish to estimate $\mathscr{C}$ from the data (X_i, Y_i).

In the majority of work in this field it is conventional to take the point process to be determined by a given number, n say, of points placed randomly into a fixed region, rather than to take it to be a Poisson process. However, in contradistinction to the case with sample size in more standard problems, the value of n will generally not be known. It seems appropriate to express this lack of knowledge by treating the number of points as a random quantity. If we specify that the integral of μ over a given, bounded region within which we are working equals 1, say, then the indeterminism of the relationship between n and μ is removed. In the context of the problem on which we are working, and in other problems of the same type, convergence rates do not depend on whether we view the number of points as Poisson with mean $n\mu$, or as n points distributed within the given region with density μ.

Next, we define the local constant estimator $\hat{g}_{\mathrm{con}}(x_0)$ at x_0. Consider the strip $\mathscr{T}(x_0) = (x_0 - h, x_0 + h) \times \mathbb{R}$, where $h > 0$ denotes the bandwidth. The local

constant estimator is obtained by locally fitting the smallest constant value that lies above all data $(X_i, Y_i) \in \mathscr{T}(x_0)$:

$$\hat{g}_{\rm con}(x_0) = \max\{Y_i : X_i \in (x_0 - h, x_0 + h)\}. \tag{2.1}$$

We estimate $\mathscr{C}$ as the curve $\hat{\mathscr{C}}_{\rm con}$ that has Cartesian representation $y = \hat{g}_{\rm con}(x)$.

If g satisfies a Lipschitz condition of order 1 then $\hat{g}_{\rm con}$ converges at rate $n^{-1/2}$, provided the bandwidth h is asymptotic to a positive constant multiple of $n^{-1/2}$. The rate $n^{-1/2}$ is minimax-optimal for such densities; see [12].

The theorem below gives an explicit expression for the limiting distribution of the local constant estimator. Let $h = C_0 n^{-1/2}$ and assume g has a continuous first derivative in a neighbourhood of x_0; call this condition $(C_{\rm con})$. (Continuity of the first derivative is needed to identify the limiting distribution.) Defining $\mu_0 = \mu\{x_0, g(x_0)\}$, $g_1 = g'(x_0)$, $\lambda = \mu_0 |g_1| C_0^2$ and $\xi = |g_1| C_0$, let F_0 denote the following distribution function, supported on $(-\infty, 1]$:

$$F_0(u) = \begin{cases} \exp(2\lambda u) & \text{if } u \leqslant -1, \\ \exp\{-\frac{1}{2}\lambda(1-u)^2\} & \text{if } -1 < u \leqslant 1, \\ 1 & \text{if } u > 1. \end{cases}$$

An additional regularity condition, for example a lower bound to the value of g, is necessary if we are to ensure the estimator is well defined with probability 1. Otherwise the value of $|\hat{g}_{\rm con}|$ can be unboundedly large. For example, if the function μ is bounded and compactly supported then the probability that there are no data pairs (X_i, Y_i) in the strip $\mathscr{T}(x_0)$ is strictly positive for each n, and in such cases $\hat{g}_{\rm con}(x_0) = -\infty$. This does not cause difficulty when describing convergence in distribution, since the probability that $\hat{g}_{\rm con}$ is well defined and finite converges to 1 as $n \to \infty$. However, when discussing convergence of moments it is a problem. Moment convergence questions must be resolved in order to determine properties of mean squared error, and hence of the optimal bandwidth.

We eliminate these difficulties by insisting that a finite, strict lower bound be placed on the value of $\hat{g}_{\rm con}$, replacing the definition of $\hat{g}_{\rm con}$ at (2.1) by the maximum of that quantity and the bound. We express the latter definition by saying that the lower bound is "reflected by" $\hat{g}_{\rm con}$. Similar assumptions will be made when we discuss higher-order polynomial estimators in Sections 3 and 4.

Theorem 1. *Assume $(C_{\rm con})$ holds. If in addition $g'(x_0) \neq 0$ then the limiting distribution of $n^{1/2}\{\hat{g}_{\rm con}(x_0) - g(x_0)\}/\xi$ is F_0; and if instead $g'(x_0) = 0$ then the limiting distribution of $-n^{1/2}\{\hat{g}_{\rm con}(x_0) - g(x_0)\}$ is exponential with mean $(2\mu_0 C_0)^{-1}$. Furthermore, if a finite strict lower bound is placed on the value of g, and is reflected by $\hat{g}_{\rm con}$, then in each case all moments converge.*

The fact that all moments converge implies in particular the convergence of mean squared error. For example, if $g'(x_0) \neq 0$ and $h = C_0 n^{-1/2}$ then

$$nE\{\hat{g}_{\rm con}(x_0) - g(x_0)\}^2 \to \alpha(C_0), \tag{2.2}$$

where $\alpha(C_0) = \xi^2 \int v^2\, dF_0(v)$. Minor changes to the proof of the theorem show that the convergence at (2.3) is uniform in $C_0 \in [C^{-1}, C]$, for any $C > 1$.

In view of the definition of F_0,

$$\alpha(u) = g_1^2 \Bigg[u^2 + (2g_1^2 u^2)^{-1} \int_{2\,|g_1|\,u^2}^{\infty} v \exp(-\mu_0\, v)\, dv - 2u^2 \int_0^1 v \exp\Big\{ -\frac{1}{2} \mu_0\, |g_1|\, u^2 (1 + v^2) \Big\} \times \big\{ \exp(\mu_0\, |g_1|\, u^2 v) - \exp(-\mu_0\, |g_1|\, u^2 v) \big\}\, dv \Bigg]. \tag{2.3}$$

Now, $\alpha(u)$ diverges to infinity as either $u \to 0$ or $u \to \infty$, and has a unique minimum at a point C_0^{opt}, say. (The latter property can be deduced on noting that if we change variable from u to $t = (|g_1|\, \mu_0)^{1/2} u$ then α is directly proportional to a function of t alone, depending on neither g_1 nor μ_0. This function can be shown numerically to have a unique minimum.) Therefore, in the case $g'(x_0) \neq 0$ the asymptotically optimal bandwidth is $h = C_0^{\mathrm{opt}} n^{-1/2}$. Fig. 1 depicts $\alpha(C_0)$ as a function of C_0 for different choices of μ_0 and g_1. We find that the minimal mean squared error is an increasing function of the slope $|g_1|$, but it decreases as the intensity μ_0 increases. The optimal C_0^{opt} decreases as $|g_1|$ or μ_0 increases.

Recall that $g_1 = g'(x_0)$. When this quantity vanishes it can be seen that $C_0^{\mathrm{opt}} = \infty$. This reflects the fact that, as may be shown more formally, when $g_1 = 0$ the asymptotically optimal bandwidth for constructing $\hat{g}_{\mathrm{con}}$ at x_0 is of larger order than $n^{-1/2}$. We shall not further consider this case.

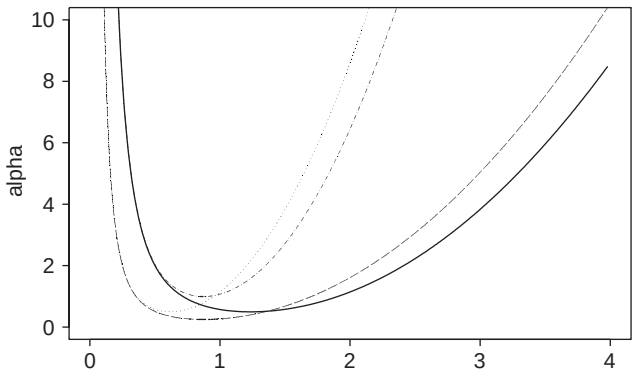

Fig. 1. Asymptotic mean squared error $\alpha(C_0)$ defined at (2.3) as a function of C_0 when $\mu_0 = 1$ and $g_1 = 1$ (solid), $\mu_0 = 2$ and $g_1 = 1$ (long-dashed), $\mu_0 = 1$ and $g_1 = 2$ (dot-and-dashed), and $\mu_0 = 2$ and $g_1 = 2$ (dotted).

3. Local linear boundary estimator

3.1. Definition of estimator

The local linear estimator $\hat{g}_{\rm lin}(x_0)$ is obtained by fitting the lowest straight line that lies above all data pairs in the strip $\mathscr{T}(x_0)$:

$$\hat{g}_{\rm lin}(x_0) = \min\{z : \text{there exists } \theta \text{ such that } Y_i \leqslant \theta(X_i - x_0) + z \text{ for all indices } i \text{ such that } X_i \in (x_0 - h, x_0 + h)\}.$$

The local linear estimator of $\mathscr{C}$ is the curve $\hat{\mathscr{C}}_{\rm lin}$ with Cartesian representation $y = \hat{g}_{\rm lin}(x)$. If g satisfies a Lipschitz condition of order 1 on its first derivative, then $\hat{g}_{\rm lin}$ has convergence rate $n^{-2/3}$ provided the bandwidth h is asymptotic to a positive constant multiple of $n^{-1/3}$. Again the rate is minimax-optimal; see [12].

An alternative way of defining a local linear estimator is to minimise not the intercept z, but the entire linear function $\theta(x - x_0) + z$: $\tilde{g}_{\rm lin}(x_0) = z_0$ where (θ_0, z_0) minimises $\sum_{i=1}^n \{\theta(X_i - x_0) + z\} I_{(x_0-h, x_0+h)}(X_i)$ with respect to (θ, z) subject to $Y_i \leqslant \theta(X_i - x_0) + z$ for all i such that $X_i \in (x_0 - h, x_0 + h)$. Limit theory can be developed for $\tilde{g}_{\rm lin}$ in much the same way we shall develop it for $\hat{g}_{\rm lin}$. Under the conditions we shall impose in Theorems 2–4 below, the two estimators have identical convergence rates. Of course, both $\hat{g}_{\rm lin}$ and $\tilde{g}_{\rm lin}$ may be generalised to estimators of higher degree, which will again share convergence rates.

3.2. Limiting distribution when $g''(x_0) < 0$

This setting, and more generally, cases where local linear methods are used to estimate a concave-downwards boundary, are unusual in the context of local polynomial boundary estimation, since bandwidth plays a relatively minor role in determining the accuracy of the estimator. Indeed, if g is strictly concave downwards on an interval $(x_0 - \delta, x_0 + \delta)$, for some $\delta > 0$, then there exists $h_1 > 0$, depending only on δ, such that, with probability 1, $\hat{g}_{\rm lin}(x_0) \leqslant g(x_0)$ for all $h \in (0, h_1]$, and $\hat{g}_{\rm lin}(x_0)$ is nondecreasing with increasing $h \in (0, h_1]$.

Nevertheless, the optimal convergence rate of $n^{-2/3}$ is attained with $h = C_1 n^{-1/3}$, for any fixed $C_1 > 0$. The constant multiple of the $n^{-2/3}$ rate can be reduced by using any bandwidth that is of strictly larger order than $n^{-1/3}$, but the rate itself is not reduced by such a choice. Therefore, when $g''(x_0) < 0$ there does not exist an asymptotically optimal bandwidth in the usual sense. The optimal constant, and the optimal rate of $n^{-2/3}$, are both attained by any bandwidth that is of strictly larger order than $n^{-1/3}$; varying the bandwidth beyond this very rudimentary prescription affects only second-order aspects of the rate, as Theorem 2 will show.

This situation is unusual, and in particular does not arise when fitting local linear estimators with $g''(x_0) > 0$. Neither does it occur when fitting higher degree local polynomial estimators, such as local quadratics or local cubics, since in such cases it is not true that the estimator is a monotone function of the bandwidth regardless of

choice of coefficients of the fitted polynomial. Nevertheless, it can sometimes happen that the optimal bandwidth for locally fitting a pth degree polynomial is an order of magnitude larger than the conventional size, $n^{-1/(p+2)}$, in particular when the true boundary can be expressed exactly by a polynomial of degree less than or equal to p.

We conclude by describing local linear estimators that are first-order optimal when $g''(x_0)<0$. Let $w<0$ and $z<0$, and define $\mathscr{L}_{wz}$, with equation $y=-(z+wz^{-1})x+w$, to be the straight line that passes through the points $(z,-z^2)$ and $(0,w)$. Note that $\mathscr{L}_{wz}$ traverses the region $\{(x,y): y\leqslant -x^2, -\infty<x<0\}$. That part of the region above $\mathscr{L}_{wz}$ is given by

$$\mathscr{S}_1(w,z)=\{(x,y): -(z+wz^{-1})x+w<y<-x^2, -\infty<x<\infty\}$$

as depicted in Fig. 2, and has area $K_1(w,z)$, given by

$$K_1(w,z)=\int_z^{w/z}\{-x^2+(z+wz^{-1})x-w\}\,dx.$$

Still assuming $w<0$ and $z<0$ we define $p_1(w,z)=\frac{1}{2}(z^2-w)$. This quantity has the following geometric interpretation: for an infinitesimal change $dz>0$ in z, with $w<0$ fixed, the area of that part of $\mathscr{S}_1(w,z)$ in the half-plane $\{(x,y): x<0\}$ decreases by $p_1(w,z)\,dz$.

Let $h=h(n)\to 0$ in such a manner that $n^{1/3}h\to\infty$. Assume g has a continuous second derivative in a neighbourhood of x_0. Call this condition (C_{lin}). (We need continuity of g'' in order to identify the limiting distribution.) Recall too the overarching conditions noted in the first paragraph of Section 2; these are assumed in Theorems 2–4 without further mention. Put $g_2=g''(x_0)$, $\kappa=\frac{1}{2}|g_2|\,\mu_0$ and

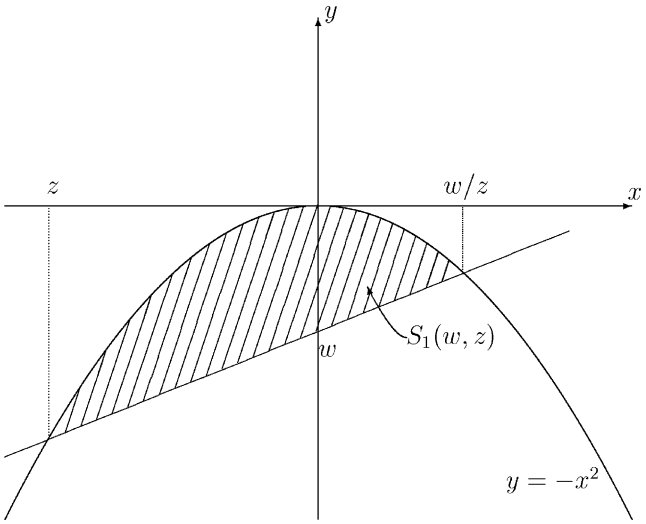

Fig. 2. The region $\mathscr{S}_1(w,z)$.

$\zeta = \frac{1}{2}|g_2|$, and define F_1 to be the distribution function that is supported on the negative half-line and defined by

$$F_1(w) = \kappa \int_{-\infty}^{0} p_1(w,z) \exp\{-\kappa K_1(w,z)\}\, dz.$$

This function can be interpreted as the distribution function of $\hat{g}_{\rm lin}$, for a strip of fixed width and computed when the support boundary is purely quadratic and the Poisson process is homogeneous below that boundary.

Theorem 2. *If* $(C_{\rm lin})$ *holds and* $g''(x_0) < 0$ *then* $n^{2/3}\{\hat{g}_{\rm lin}(x_0) - g(x_0)\}/\zeta$ *has limiting distribution* F_1. *Furthermore, if a finite strict lower bound is placed on the value of* g, *and is reflected by* $\hat{g}_{\rm lin}$, *then all moments converge.*

Reflecting the fact that first-order properties of $\hat{g}_{\rm lin}(x_0)$ do not depend on h if $n^{1/3}h \to \infty$, the following may also be proved. Let h_1, h_2 denote bandwidth sequences (deterministic functions of n) which satisfy $h_j \to 0$ and $n^{1/3}h_j \to \infty$ as $n \to \infty$, and write $\hat{g}_{\rm lin}(\cdot|h_j)$ for the corresponding local linear estimator. Then the probability that $\hat{g}_{\rm lin}(x_0|h_1) = \hat{g}_{\rm lin}(x_0|h_2)$ converges to 1 as $n \to \infty$.

3.3. Limiting distribution when $g''(x_0) > 0$

In the present setting the limiting distribution, after rescaling, is supported on $(-\infty, 1]$ instead of $(-\infty, 0]$ (where it was supported when $g''(x_0) < 0$). If $w < 0$, put $z_0 = \min(1, \sqrt{-w})$. Define the set

$$\mathscr{S}_2(w,z) = \{(x,y)\colon (z - wz^{-1})x + w < y < x^2, -1 < x < 1\},$$

for either $0 \leqslant w < 1$ and $-1 < z < 0$, or $w < 0$ and $-z_0 < z < z_0$. The area of $\mathscr{S}_2(w,z)$ is given by

$$K_2(w,z) = \int_{(-1,z)\cup(\min(-w/z,1),1)} \{x^2 - (z - wz^{-1})x - w\}\, dx$$

if $0 \leqslant w < 1$ and $-1 < z < 0$, by

$$K_2(w,z) = \int_{(-1,z_1)\cup(z,1)} \{x^2 - (z - wz^{-1})x - w\}\, dx$$

if $w < 0$ and $-z_0 < z < 0$, and by

$$K_2(w,z) = \int_{(-1,z)\cup(z_2,1)} \{x^2 - (z - wz^{-1})x - w\}\, dx$$

if $w < 0$ and $0 < z < z_0$, where $z_1 = \max(-1, -w/z)$ and $z_2 = \min(-w/z, 1)$.

Let $h = n^{-1/3}C_1$ denote the bandwidth used to construct the estimator introduced in Section 3.1. Let κ and ζ be as in Section 3.2, define

$$p_2(w,z) = \begin{cases} \frac{1}{2}(z^2+w)(z^{-2}-1) & \text{if } 0<w<1 \text{ and } -1<z<0, \\ -\frac{1}{2}(1+wz^{-2}) & \text{if } w<0 \text{ and } 0<z<z_0, \\ -\frac{1}{2}(z^2+w) & \text{if } \max(w,-1)<z<0, \\ -\frac{1}{2}(z^2+w)(1+z^{-2}-w^2z^{-4}) & \text{if } -z_0<z<\max(w,-1)<0 \end{cases}$$

and put

$$F_2(w,C_1) = \begin{cases} \kappa C_1^3 \int_{-1}^{0} p_2(w,z)\exp\{-\kappa C_1^3 K_2(w,z)\}\,dz & \text{if } 0<w<1, \\ \kappa C_1^3 (z_0 - wz_0^{-1})\exp\{-\frac{1}{3}\kappa C_1^3(2-3w)\} & \\ \quad +\kappa C_1^3 \int_{-z_0}^{z_0} p_2(w,z)\exp\{-\kappa C_1^3 K_2(w,z)\}\,dz & \text{if } w<0. \end{cases}$$

The function p_2 has a geometric interpretation: if $0<w<1$ and $-1<z<0$, or if $w<0$ and $-z_0<z<z_0$, then $p_2(w,z)\,dz$ equals the absolute value of the change in the area of $\mathscr{S}_2(w,z)$, restricted to the half-plane $\{(x,y): x<0\}$, for an infinitesimal change dz in z.

Theorem 3. *If* $(C_{\rm lin})$ *holds and* $g''(x_0)>0$ *then* $n^{2/3}\{\hat g_{\rm lin}(x_0) - g(x_0)\}/(\zeta C_1^2)$ *has limiting distribution* $F_2(\cdot, C_1)$. *Furthermore, if a finite strict lower bound is placed on the value of* g, *and is reflected by* $\hat g_{\rm lin}$, *then all moments converge.*

The fact that moments converge implies that

$$n^{4/3}\,E\{\hat g_{\rm lin}(x_0) - g(x_0)\}^2 \to \beta(C_1) \equiv \zeta^2 C_1^4 \int v^2\,dF_2(v,C_1). \tag{3.1}$$

Analogously to α defined at (2.3), β is a positive function on the positive half-line, and $\beta(u)$ diverges to infinity as either $u\to 0$ or $u\to\infty$, its minimum occurring at a

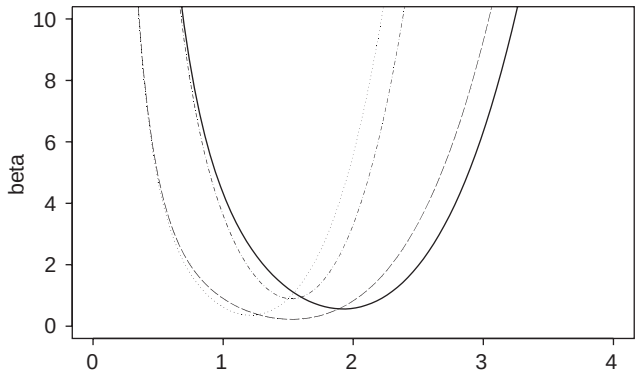

Fig. 3. Asymptotic mean squared error $\beta(C_1)$ defined at (3.1) as a function of C_1 when $\mu_0 = 1$ and $g_2 = 1$ (solid), $\mu_0 = 2$ and $g_2 = 1$ (long-dashed), $\mu_0 = 1$ and $g_2 = 2$ (dot-and-dashed), and $\mu_0 = 2$ and $g_2 = 2$ (dotted).

unique point $C_1^{\rm opt}$, say. The asymptotically optimal bandwidth is thus $h = C_1^{\rm opt} n^{-1/3}$. As in the local constant case, β is a particularly complicated function and $C_1^{\rm opt}$ cannot be determined analytically. Fig. 3 depicts $\beta(C_1)$ as a function of C_1 for different choices of μ_0 and g_2. Similarly to Fig. 1, it shows that the minimal mean squared error is an increasing (decreasing) function of the curvature g_2 (the intensity μ_0), and that the optimal $C_1^{\rm opt}$ decreases as g_2 or μ_0 increases.

3.4. Limiting distribution when $g''(x_0) = 0$

This case is analogous, in the local linear setting, to the context $g'(x_0)$ mentioned for local constant estimators at the end of Section 2. Optimal choice of bandwidth depends on high-order derivatives of g in the neighbourhood of x_0, and convergence rates faster than $n^{-2/3}$ can be achieved using bandwidths of larger order than $n^{-1/3}$. However, for the sake of completeness we discuss below the properties of $\hat g_{\rm lin}$ when it is constructed using bandwidth $h = C_1 n^{-1/3}$.

We begin by describing the limiting distribution. The area $K_3(w,s)$ of the region

$$\mathscr{S}_3(w,s) = \{(x,y)\colon sx + w < y < 0, \quad -1 < x < 1\},$$

$= -\frac{1}{2}s(1 + ws^{-1})^2$ if $s < w$, $-2w$ if $w < s < -w$, and $\frac{1}{2}s(1 - ws^{-1})^2$ if $s > -w$. Define $p_3(w,s) = w^2/2s^2$ if $s < w$, and $\frac{1}{2}$ otherwise. We may interpret $p_3(w,s)\,ds$ as the absolute value of the change in the area of the intersection of $\mathscr{S}_3(w,z)$ with the half-plane $\{(x,y) : x < 0\}$, for an infinitesimal slope change ds. Let F_3 denote the distribution supported on the negative half-line and defined there by

$$F_3(w, C_1) = \mu_0 C_1 \int_{-\infty}^{\infty} p_3(w,z) \exp\{-\mu_0 C_1 K_3(w,z)\}\, dz.$$

Theorem 4. *If* $(C_{\rm lin})$ *holds and* $g''(x_0) = 0$ *then* $n^{2/3}\{\hat g_{\rm lin}(x_0) - g(x_0)\}$ *has limiting distribution* $F_3(\cdot, C_1)$. *Furthermore, if a finite strict lower bound is placed on the value of* g, *and reflected by* $\hat g_{\rm lin}$, *then all moments converge.*

Analogues of Theorems 1–4 can be proved for general pth degree local polynomial estimators, fitted to boundaries that have $p+1$ continuous derivatives. The optimal bandwidth is generally of size $n^{-1/(p+2)}$, and the convergence rate is $n^{-(p+1)/(p+2)}$.

4. Bandwidth selection

4.1. A bootstrap method

In conventional curve estimation problems there is a variety of techniques for choosing bandwidth. They include plug-in rules, cross-validation and the bootstrap. In the present context plug-in rules are unattractive since, even in the relatively simple case of local constant smoothing, no explicit expression is available for the multiplicative constant in the formula for the asymptotically optimal bandwidth.

See, for example, the discussion centring on formulae (2.3) and (3.1). Cross-validation, too, is difficult in the present case since, in spatial or multivariate problems, there is no elementary surrogate for the boundary when calculating the cross-product contribution to mean squared error. We suggest instead a bootstrap algorithm, based on a new method for simulating the Poisson process $\mathscr{P}$ beneath an estimate of the boundary curve.

We shall describe methods and properties for general pth degree polynomial smoothers, defined by

$$\hat{g}(x_0) = \min\Bigg\{ z : \text{ there exist } \theta_1, \ldots, \theta_p \text{ such that } Y_i \leqslant z + \sum_{1 \leqslant j \leqslant p} \theta_j (X_i - x_0)^j \text{ for all } i \text{ such that } X_i \in (x_0 - h, x_0 + h) \Bigg\}. \tag{4.1}$$

First we construct a pilot estimator, $\hat{g}_0$ say, using a bandwidth h_0 in place of h above. Then we oversmooth $\hat{g}_0$, using a kernel approach with bandwidth $h_1 > h_0$, obtaining a second pilot estimator $\hat{g}_{\text{pil}}$:

$$\hat{g}_{\text{pil}}(x) = \frac{1}{h_1} \int \hat{g}_0(y) K\left(\frac{x-y}{h_1}\right) dy. \tag{4.2}$$

The corresponding pilot boundary-curve estimator, $\hat{\mathscr{C}}_{\text{pil}}$, is the curve with Cartesian representation $y = \hat{g}_{\text{pil}}(x)$.

We simulate $\mathscr{P}$ below $\hat{\mathscr{C}}_{\text{pil}}$, using a k-nearest-neighbour method where $k \geqslant 2$, and generating a new, bootstrap point process $\mathscr{P}^*$. In outline, the method involves the following four steps. (In each case, motivation and further information is given in parentheses.) (a) Remove all points that lie above $\hat{\mathscr{C}}_{\text{pil}}$. (Since $\hat{\mathscr{C}}_{\text{pil}}$ is so close to the real boundary then there is little reliable information about the intensity above $\hat{\mathscr{C}}_{\text{pil}}$, and so we should not rely on data there.) (b) Generate "pseudodata" above $\hat{\mathscr{C}}_{\text{pil}}$ by reflecting points below $\hat{\mathscr{C}}_{\text{pil}}$ in that curve. (We generate these pseudodata so as to replace the data we have removed. The intensity of the pseudodata is known reliably, since we have good information about point-process intensity below $\hat{\mathscr{C}}_{\text{pil}}$. We need the pseudodata so as to not suffer edge effects when we produce the bootstrap points.) (c) Generate bootstrap points, both above and below $\hat{\mathscr{C}}_{\text{pil}}$, by distributing Poisson numbers of points in circular neighbourhoods of each real point below $\hat{\mathscr{C}}_{\text{pil}}$ and each pseudodata point above $\hat{\mathscr{C}}_{\text{pil}}$. (We base this step of the algorithm on Poisson-distributed points within discs, but an alternative approach would be to construct a Dirichlet tessellation and use the cells of this instead of discs.) (d) Remove the bootstrap points generated above $\hat{\mathscr{C}}_{\text{pil}}$, so that the bootstrap points that remain have $\hat{\mathscr{C}}_{\text{pil}}$ as the edge of the support of their intensity. (This step is of course

necessary if, in the bootstrap world, the pilot boundary estimator is to truly represent the boundary of our simulated data.)

In step (c), the disc-based method for generating bootstrap data in circular neighbourhoods of each real point, or each pseudodatum, can be thought of as a device for capturing the local variation of the original point process without having to actually estimate its intensity. Doing the latter would require selecting another bandwidth, and we felt that should be avoided where possible.

Next we give detail; steps (a)–(d) below correspond directly to their counterparts in the outline above. (a) Delete each point (X_i, Y_i) of $\mathscr{P}$ that lies above $\hat{\mathscr{C}}_{\rm pil}$. Let $\mathscr{Z}_1$ denote the data that remain (i.e. all the data below $\hat{\mathscr{C}}_{\rm pil}$), and suppose $\mathscr{Z}_1$ contains m distinct points. (b) Impute m pseudodata above $\hat{\mathscr{C}}_{\rm pil}$, by placing into the plane the point $(X_i, 2\hat{g}_{\rm pil}(X_i) - Y_i)$ for each $(X_i, Y_i) \in \mathscr{Z}_1$. Let $\mathscr{Z}$ denote the union of the m data in $\mathscr{Z}_1$ and the m pseudodata. For each $Z \in \mathscr{Z}$, compute the distance $d(Z)$ from Z to the kth nearest point in $\mathscr{Z}$. Let $\mathscr{D}(Z)$ be the disc centred at Z and with radius $d(Z)$; and conditional on $\mathscr{P}$, generate the set $\mathscr{N} = \{N(Z), Z \in \mathscr{Z}\}$ of mutually independent Poisson-distributed random variables, each with unit mean. (c) Conditional on $\mathscr{P} \cup \mathscr{N}$, generate mutually independent random variables $Z_i^*(Z)$, for $1 \leqslant i \leqslant N(Z)$ and $Z \in \mathscr{Z}$, in such a way that for each $Z \in \mathscr{Z}$ the random variables $Z_i^*(Z)$, $1 \leqslant i \leqslant N(Z)$ are uniformly distributed over $\mathscr{D}(Z)$. Write $\mathscr{P}_1^*$ for the set of all the $Z_i^*(Z)$'s. (d) Let $\mathscr{P}^*$ be the set of points in $\mathscr{P}_1^*$ that lie below $\hat{\mathscr{C}}_{\rm pil}$.

Using $\mathscr{P}^*$ rather than $\mathscr{P}$, and bandwidth h, calculate the analogue $\hat{g}^*$ of $\hat{g}$, the latter defined at (4.1). Compute the bootstrap estimator,

$$\widehat{\mathrm{MSE}}(x,h) = E[\{\hat{g}^*(x) - \hat{g}_{\rm pil}(x)\}^2 | \mathscr{P}],$$

of the mean squared error, $\mathrm{MSE}(x,h) = E[\{\hat{g}(x) - g(x)\}^2]$, of $\hat{g}$. To compute a local empirical bandwidth $\hat{h}_{\rm loc}(x_0)$ for estimating g at a particular point x_0, or a global empirical bandwidth $\hat{h}_{\rm glob}$ for estimating $g(x)$ for x in an interval $\mathscr{I}$, put

$$\hat{h}_{\rm loc}(x_0) = \arg\min_h \widehat{\mathrm{MSE}}(x_0,h) \quad \text{and} \quad \hat{h}_{\rm glob} = \arg\min_h \int_{\mathscr{I}} \widehat{\mathrm{MSE}}(x,h)\,dx, \tag{4.3}$$

respectively. We may consider $\hat{h}_{\rm loc}$ and $\hat{h}_{\rm glob}$ to be respective approximations to

$$h_{\rm loc}(x_0) = \arg\min_h \mathrm{MSE}(x_0,h) \quad \text{and} \quad h_{\rm glob} = \arg\min_h \int_{\mathscr{I}} \mathrm{MSE}(x,h)\,dx. \tag{4.4}$$

We shall assume that $h_{\rm loc}$ and $h_{\rm glob}$ are both of size $n^{-1/(p+2)}$, this being the order of bandwidth that in most instances minimises $\mathrm{MSE}(x,h)$. In the case of $h_{\rm loc}$, and for local constant and local linear estimators, this follows directly from the convergence-of-moments properties in Theorems 1 and 3; see in particular (2.2), (2.3) and (3.1), and the discussions of those properties. Results for higher-order local polynomial fits are similar. So too is the case of $h_{\rm glob}$, provided we impose a regularity condition such as $(C_{\rm bw})$ below.

4.2. Asymptotic optimality

The empirical bandwidths $\hat{h}_{\rm loc}$ and $\hat{h}_{\rm glob}$ are likewise asymptotic to their optimal, theoretical counterparts, in the senses described by the theorem below. To obtain that result, when generating the points in $\mathscr{P}^*$ we assume $k = k(n)$ (the near-neighbour index) is an integer satisfying $n^{\varepsilon} \leqslant k \leqslant n^{1-\varepsilon}$ for some $\varepsilon \in (0, \frac{1}{2})$. When calculating $\hat{g}_0$ and $\hat{g}_{\rm pil}$ we use bandwidths $h_0 \asymp n^{-1/(p+2)}$ and h_1 satisfying $n^{(1-\varepsilon)/(p+2)} h_1 \to \infty$ and $n^{(1+\varepsilon)/2(p+2)} h_1 \to 0$ for some $\varepsilon > 0$, and we employ a compactly supported kernel K that has $p+1$ continuous derivatives and satisfies $\int K = 1$ and $\int u^j K(u)\, du = 0$ for $1 \leqslant j \leqslant p$. (Note that $a_n \asymp b_n$, for positive sequences a_n and b_n, means that a_n/b_n is bounded away from zero and infinity.)

Assume too that there exists an open set $\mathscr{J}$, containing x_0, such that $g^{(p+1)}$ exists, is continuous and does not vanish in $\mathscr{J}$, that $h_{\rm loc}, h_{\rm glob} \asymp n^{-1/(p+2)}$, and that the compact interval $\mathscr{I}$ in the definitions at (4.3) and (4.4) is contained within $\mathscr{J}$. Suppose the Poisson intensity $\mu(x, y)$ vanishes for $y > g(x)$, and that for some $\varepsilon > 0$, and within the region $\{(x, y)\colon x \in \mathscr{J} \text{ and } g(x) - \varepsilon \leqslant y \leqslant g(x)\}$, μ is both bounded away from 0 and continuous. Finally, assume that a finite lower bound is placed on the value of g, so that estimator and its bootstrap version are replaced by the maximum of the bound and the estimator's traditional form. Denote by $(C_{\rm bw})$ the union of the conditions in this and the previous paragraph; "bw" denotes "bandwidth".

The assumption that $g^{(p+1)}$ does not vanish is unnecessary if we restrict attention to the case of global bandwidth choice; there it is necessary only to assume $g^{(p+1)}$ is nonvanishing on a nondegenerate subinterval of $\mathscr{I}$. Furthermore, the assumption that $h_0 \asymp n^{-1/(p+2)}$, imposed in $(C_{\rm bw})$, may be relaxed. It implies that the pilot bandwidth h_0 is of optimal size, although h_0 might not involve the optimal value of the multiplicative constant. However, in practice an empirical version of h_0, satisfying $h_0 \asymp n^{-1/(p+2)}$, would be assured by iterating the suggested approach, and so we have not bothered to relax the condition.

Theorem 5. *If* $(C_{\rm bw})$ *holds then* $\hat{h}_{\rm loc}(x_0)/h_{\rm loc}(x_0) \to 1$ *and* $\hat{h}_{\rm glob}/h_{\rm glob} \to 1$ *in probability.*

4.3. Numerical study

We present the results of a numerical experiment demonstrating the effectiveness of the bootstrap bandwidth selection method for local linear estimators. We considered a Poisson process with constant intensity $n\mu(\cdot) \equiv 100$ on the region under the boundary $y = x^2$. First, we simulated 5000 datasets from the Poisson process to find the local optimal bandwidth $h_{\rm loc}(0)$ in this setting. For each dataset we computed the local linear estimate at $x_0 = 0$. We approximated the mean squared error using these 5000 estimates. The mean squared error, as a function of the bandwidth h, is depicted as the solid curve in Fig. 4, where the vertical line indicates the location of $h_{\rm loc}(0)$; we found that the latter quantity was $h_{\rm loc}(0) = 0.2795$.

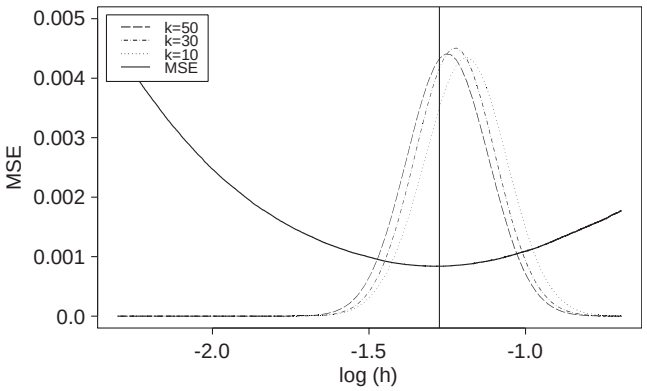

Fig. 4. Mean squared error of $\hat{g}_{\mathrm{lin}}(0)$ as a function of $\log h$ (solid curve), and kernel density estimates of the distributions of $\log \hat{h}_{\mathrm{loc}}(0)$ when $k = 10$ (dotted curve), $k = 30$ (dot-and-dashed curve), and $k = 50$ (long-dashed curve). The vertical line indicates the location of the optimal local bandwidth, $\log h(0)$.

To see how well the bootstrap bandwidth selector $\hat{h}_{\mathrm{loc}}(0)$ estimated the local optimum $h_{\mathrm{loc}}(0)$, we simulated 100 datasets from the Poisson process. For the pilot estimator of g, we used the local linear smoother $\hat{g}_{\mathrm{lin}}$ with bandwidth $h_0 = h_{\mathrm{loc}}(0)$. For the second smoothed pilot estimator $\hat{g}_{\mathrm{pil}}$, defined at (4.2), we took $h_1 = 1.5h_{\mathrm{loc}}(0)$ and $K(u) = I_{[-1/2,1/2]}(u)$. For each dataset, 100 bootstrap samples $\mathscr{P}^*$ were generated to approximate the bootstrap estimate of mean squared error. Kernel density estimates of the distributions of $\log \hat{h}_{\mathrm{loc}}(0)$ for $k = 10$, 30 and 50 are overlaid in Fig. 4 as the dotted, dot-and-dashed and long-dashed curves, respectively. There are notable improvements as we move from $k = 10$ to 50. We tried other values of k in the range from 2 to 60, and found that the mean squared error of the bootstrap bandwidth selector is least at $k = 50$. The relative mean squared errors $E[\{\hat{h}_{\mathrm{loc}}(0)/h_{\mathrm{loc}}(0)\} - 1]^2$ for $k = 10$, 30 and 50 were 0.0166, 0.0096 and 0.0079, respectively.

Fig. 5 shows the resulting estimates $\hat{g}_{\mathrm{lin}}(x)$ of the boundary $g(x) = x^2$ which used the bootstrap bandwidth selector $\hat{h}_{\mathrm{loc}}(x)$. We depicted the boundary estimates for ten simulated datasets from a Poisson process with the constant intensity 100 on the region under the boundary $y = x^2$. For each dataset, we computed $\hat{h}_{\mathrm{loc}}(x)$ for 20 equally spaced points x in the interval $[-1, 1]$ in the same way as described in the above paragraph. We calculated $\hat{g}_{\mathrm{lin}}(x)$ with the obtained bootstrap bandwidths for these values of x, and depicted the whole curve on the interval $[-1, 1]$ by linear interpolation. The two panels of Fig. 5 correspond to the cases where $k = 10$ and 50 are chosen. Both panels exhibit quite good performance of the resulting boundary estimates. Comparing the two panels we see that $k = 50$ yields better estimates than $k = 10$, especially in the central area around $x = 0$.

We computed the bootstrap bandwidths using other values of h_1 in the range from $h_1 = 0$ (i.e., no pre-smoothing) to $h_1 = 3\,h_{\mathrm{loc}}(0)$. Also, we investigated the case where the Poisson process has intensity $n\mu(\cdot) \equiv 50$. We observed that taking large (small) h_1

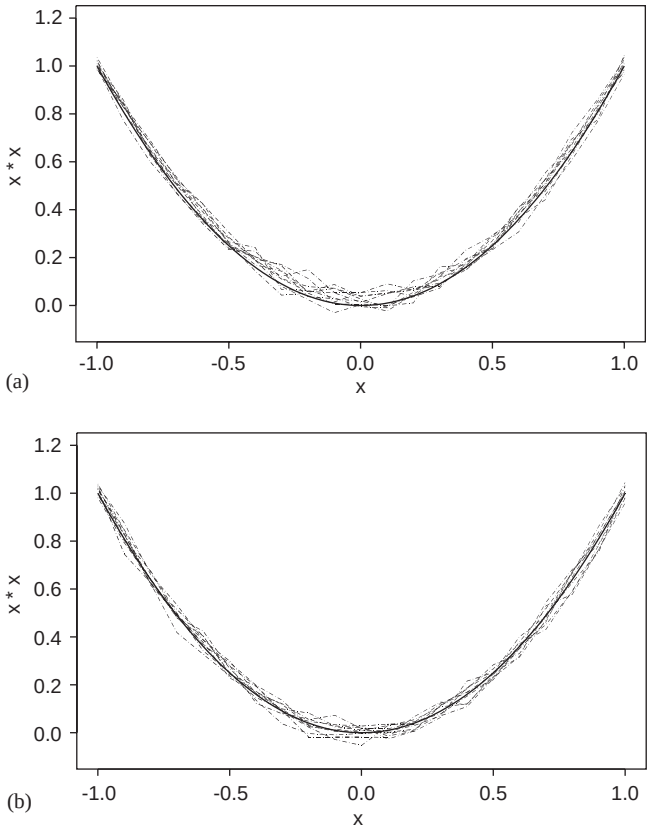

Fig. 5. Local linear boundary estimates for ten simulated datasets from a Poisson process with the constant intensity 100 on the region under the boundary $y = x^2$. The panel (a) corresponds to the case where $k = 10$ was used in the bootstrap bandwidth selection, and the panel (b) shows the results when $k = 50$ was used.

decreases (increases) the variability, $\mathrm{var}\{\hat{h}_{\mathrm{loc}}(0)/h_{\mathrm{loc}}(0)\}$, but produces large (small) bias, $E[\{\hat{h}_{\mathrm{loc}}(0)/h_{\mathrm{loc}}(0)\} - 1]$, respectively. The results showed that the overall relative mean squared error of $\hat{h}_{\mathrm{loc}}$ is least for $h_1 \in (h_{\mathrm{loc}}(0), 1.5\, h_{\mathrm{loc}}(0))$.

5. Technical details

We omit the proofs of Theorems 1 and 4, since they are similar to but less involved than those of Theorems 2 and 3.

5.1. *Proof of Theorem 2*

Without essential loss of generality, we may assume that the boundary is exactly quadratic. Consider the linear transformation that takes (X_i, Y_i) to (X_i', Y_i'), where

$X_i' = n^{1/3}(X_i - x_0)$ and $Y_i' = n^{2/3}\{Y_i - g_0 - g_1(X_i - x_0)\}$, in which $g_0 = g(x_0)$ and $g_1 = g'(x_0)$. In the new coordinate system the support of $\mathscr{P}$ has as its boundary the curve $\mathscr{C}$ with equation $y = \frac{1}{2} g_2 x^2$, and has intensity function μ_n say, where μ_n is a bounded, continuous function in the half-plane below the boundary and, for each sequence $\varepsilon_n \downarrow 0$, satisfies

$$\sup' |\mu_n(x, y) - \mu_0| \to 0, \tag{5.1}$$

where $\sup'$ denotes the supremum over pairs (x, y) such that $|x| \leqslant \varepsilon_n n^{1/3}$ and $-\varepsilon_n n^{2/3} \leqslant y \leqslant \frac{1}{2} g_2 x^2$.

Let $\mathscr{P}_1$ denote the Poisson process with uniform intensity μ_0 in the region below $\mathscr{C}$, and let V_1 be the version of $\hat{g}_{\text{lin}}$ that is obtained using these data. Then, noting (5.1), it may be proved from properties of Poisson processes that the distribution of V_1 is the limiting distribution of $n^{2/3}\{\hat{g}_{\text{lin}}(x_0) - g(x_0)\}$. By stretching the vertical axis by the factor $\zeta^{-1} = -2/g_2$, we obtain another new Poisson process $\mathscr{P}_2$ with uniform intensity $\kappa = -\frac{1}{2} g_2 \mu_0$ under the curve given by the equation $y = -x^2$. (The problem at hand is now that of estimating "$g(x_0) = 0$" at "$x_0 = 0$".) Let W_1 denote the version of $\hat{g}_{\text{lin}}$ that is obtained using the data $\mathscr{P}_2$. Then W_1 has the same distribution as V_1/ζ. We shall complete the proof of Theorem 2 by showing that W_1 has distribution F_1.

Recall that for given $w < 0$ and $z < 0$, $\mathscr{L}_{wz}$ denotes the straight line passing through $(z, -z^2)$ and $(0, w)$. Define

$$\begin{aligned} Z_1 = \max\{z < 0 &: \mathscr{L}_{wz} \text{ contains at least} \\ &\text{one point } (X_i, Y_i) \text{ satisfying } -\infty < X_i < 0\}. \end{aligned} \tag{5.2}$$

In a slight abuse of notation we shall refer to the Poisson points in $\mathscr{P}_2$ as (X_i, Y_i). We shall show that for an infinitesimal change $dz > 0$,

$$P\{W_1 \leqslant w, Z_1 \in (z, z + dz)\} = \kappa p_1(w, z) \exp\{-\kappa K_1(w, z)\}\, dz, \tag{5.3}$$

uniformly in z in compact subsets of $(-\infty, 0]$. The theorem follows if we establish (5.3).

Consider moving z in the negative direction, along the negative half-line, starting from the origin, and stopping when the line $\mathscr{L}_{wz}$ hits (for the first time) a data point in $\mathscr{P}_2$. Then Z_1 equals the value of z when this motion ceases. Thus, by definition of Z_1, that part of $\mathscr{S}_1(w, Z_1)$ restricted to the half-plane $\mathscr{H} = \{(x, y): x < 0\}$ contains no data. Furthermore, it can be seen that $W_1 \leqslant w$ if and only if $\mathscr{S}_1(w, Z_1)$ contains no data. It may be shown from these results, using properties of Poisson processes, that the probability on the left-hand side of (5.3) equals the probability $Q(w, z, dz)$ that there exists exactly one point in the set $\{\mathscr{S}_1(w, z) \triangle \mathscr{S}_1(w, z + dz)\} \cap \mathscr{H}$ and no other point in $\mathscr{S}_1(w, z)$. (Here, $A \triangle B$ denotes the symmetric difference of sets A and B.) Note that $p_1(w, z)\, dz$ and $K_1(w, z)$, defined in Section 5.1, are the areas of the sets $\{\mathscr{S}_1(w, z) \triangle \mathscr{S}(w, z + dz)\} \cap \mathscr{H}$ and $\mathscr{S}_1(w, z)$, respectively. This shows that W_1 has distribution F_1, and completes the proof of the convergence-in-distribution part of Theorem 2.

To establish convergence of moments we note first that the estimator $\hat{g}_{\rm lin}$ increases monotonically as the dataset $\mathscr{P}$ grows. That is, if we expand the dataset $\mathscr{P}$ to $\mathscr{P}^\dagger$, where $\mathscr{P} \subseteq \mathscr{P}^\dagger$, then the respective local linear estimators $\hat{g}_{\rm lin}$ and $\hat{g}^\dagger_{\rm lin}$ satisfy $\hat{g}_{\rm lin} \leqslant \hat{g}^\dagger_{\rm lin}$. Call this the "monotonicity property". Observe too that, given a bandwidth h, we may delete all data outside the box $\mathscr{B} = (x_0 - h, x_0 + h) \times (g(x_0) - h^{1/2}, g(x_0) + h^{1/2})$ and, with probability converging to 1 at an exponentially fast rate, the estimator $\hat{g}_{\rm lin}$ will remain unchanged. Since this convergence rate is so fast, and since a finite lower bound has been placed on the value of the estimator, then if we were to replace (the lower-bounded version of) $\hat{g}_{\rm lin}$ by an estimator constructed solely from data within $\mathscr{B}$ (and reflecting the lower bound), its moments would differ from those of $\hat{g}_{\rm lin}$ in terms that are of smaller order than the inverse of any polynomial in n.

Therefore it suffices to prove convergence of moments in the case where $\hat{g}_{\rm lin}$ has been constructed solely from data within $\mathscr{B}$, and reflecting the lower bound. In a slight abuse of notation we shall write μ for the function that equals the original μ on $\mathscr{B}$ and which vanishes elsewhere, we shall write $\mathscr{P}$ for a dataset generated by a Poisson process with intensity $n\mu$ for this new μ, and we shall let $\hat{g}_{\rm lin}$ denote the corresponding local linear estimator.

Introduce superscripts L and U to denote "lower" and "upper", respectively, and construct new boundary curves $\mathscr{C}^{\rm L}$ and $\mathscr{C}^{\rm U}$, determined by respective quadratic equations $y = g(x_0) + a^{\rm L} + (x - x_0)\, g'(x_0) + \frac{1}{2}(x - x_0)^2\, g''(x_0)$ and $y = g(x_0) + a^{\rm U} + (x - x_0)\, g'(x_0) + \frac{1}{2}(x - x_0)^2\, g''(x_0)$. Here, $a^{\rm L}$ and $a^{\rm U}$ are deterministic and depend on n. They have the properties

$$a^{\rm L} < 0 < a^{\rm U} \quad \text{and} \quad a^{\rm U} - a^{\rm L} = o(h^2), \tag{5.4}$$

and are chosen such that $\mathscr{C}$ is contained strictly between $\mathscr{C}^{\rm L}$ and $\mathscr{C}^{\rm U}$ within the strip $(x_0 - h, x_0 + h) \times \mathbb{R}$.

Let $\mu^{\rm L}$ (respectively, $\mu^{\rm U}$) denote the function that, within that part of $\mathscr{B}$ below $\mathscr{C}^{\rm L}$ (below $\mathscr{C}^{\rm U}$), equals the infimum [supremum] of μ on $\mathscr{B}$, and which vanishes above the respective boundary curve and also vanishes outside $\mathscr{B}$. Noting that $\mathscr{P}$ now consists solely of data in $\mathscr{B}$, we see that we may write $\mathscr{P} = \mathscr{P}^{\rm L} \cup \mathscr{Q}^{\rm L}$ and $\mathscr{P}^{\rm U} = \mathscr{P} \cup \mathscr{Q}^{\rm U}$, where $\mathscr{P}^{\rm L}$, $\mathscr{Q}^{\rm L}$ and $\mathscr{Q}^{\rm U}$ are datasets generated by independent Poisson processes with respective intensities $n\mu^{\rm L}$, $n(\mu - \mu^{\rm L})$ and $n(\mu^{\rm U} - \mu)$; $\mathscr{P}^{\rm U}$ is a Poisson process with intensity $n\mu^{\rm U}$; and $\mathscr{P}^{\rm L}$ and $\mathscr{P}^{\rm U}$ vanish above $\mathscr{C}^{\rm L}$ and $\mathscr{C}^{\rm U}$, respectively.

Let $\hat{g}^{\rm L}_{\rm lin}$ and $\hat{g}^{\rm U}_{\rm lin}$ denote the versions $\hat{g}_{\rm lin}$ computed from $\mathscr{P}^{\rm L}$ and $\mathscr{P}^{\rm U}$, respectively, instead of $\mathscr{P}$. In view of the monotonicity property we have $\hat{g}^{\rm L}_{\rm lin} \leqslant \hat{g}_{\rm lin} \leqslant \hat{g}^{\rm U}_{\rm lin}$. Therefore, defining $\delta^{\rm L} = n^{2/3}\, a^{\rm L}/\zeta$, $\delta^{\rm U} = n^{2/3}\, a^{\rm U}/\zeta$,

$$V^{\rm L} = n^{2/3}\{\hat{g}^{\rm L}_{\rm lin}(x_0) - g(x_0) - a^{\rm L}\}/\zeta, \quad V^{\rm U} = n^{2/3}\{\hat{g}^{\rm U}_{\rm lin}(x_0) - g(x_0) - a^{\rm U}\}/\zeta,$$

$$V = n^{2/3}\{\hat{g}_{\rm lin}(x_0) - g(x_0)\}/\zeta,$$

we have

$$V^{\rm L} + \delta^{\rm L} \leqslant V \leqslant V^{\rm U} + \delta^{\rm U}. \tag{5.5}$$

Note that $\hat{g}_{\rm lin}^{\rm L}$ and $\hat{g}_{\rm lin}^{\rm U}$ are computed from data generated by Poisson processes that are perfectly homogeneous below perfectly quadratic boundaries. However, on the present occasion the intensities are not supported outside $\mathscr{B}$, and in particular are not supported a distance $O(h^{1/2})$ below their respective boundaries $\mathscr{C}^{\rm L}$ and $\mathscr{C}^{\rm U}$; see the definition of $\mathscr{B}$ five paragraphs above. If instead we extend the Poisson processes in a homogeneous manner, so that they have respective intensities $n\mu^{\rm L}$ and $n\mu^{\rm U}$ in the semi-infinite strips below $\mathscr{C}^{\rm L}$ and $\mathscr{C}^{\rm U}$, leading to redefinitions of $V^{\rm L}$ and $V^{\rm U}$; then the distributions of $V^{\rm L}$ and $V^{\rm U}$ can be written down exactly. They are in fact no more than slight reparametrizations of the distribution of W_1, which is F_1. In this case it is trivial to prove that the moments of $V^{\rm L}$ and $V^{\rm U}$ converge to those of W_1. Returning to the original definitions of $V^{\rm L}$ and $V^{\rm U}$, for which (5.5) holds, and noting that the strict lower bound for $g(x_0)$ is reflected in the definition of $\hat{g}_{\rm lin}$, we see that the change in the definitions impacts any one of the moments by an amount that is of smaller order than the inverse of any polynomial in n. This result, (5.4) and (5.5) imply that the moments of V converge to those of W_1, as had to be shown.

5.2. *Proof of Theorem 3*

The first half of the proof is similar to that of Theorem 2. Here, g_2 plays the role of $-g_2$ there, V_1 is replaced by the version that is obtained now using the data in the strip $(-C_1, C_1) \times \mathbb{R}$, and the horizontal and the vertical axes are stretched by the factors C_1^{-1} and $(\zeta C_1^2)^{-1}$, respectively. Thus it can be seen that $n^{2/3}\{\hat{g}_{\rm lin}(x_0) - g(x_0)\}/(\zeta C_1^2)$ converges in distribution to W_2, equal to the version of $\hat{g}_{\rm lin}$ that is computed from Poisson data having uniform intensity κC_1^3 within the set $\{(x,y): y \leqslant x^2,\ -1 < x < 1\}$. Note that $P(W_2 \geqslant 1) = 0$. The case $0 < w < 1$ can be dealt with similarly to the proof of Theorem 2, where instead of Z_1 at (5.2) we need to define Z_2 to be the least value of $z < 0$ such that the line passing through (z, z^2) and $(0, w)$, still denoted by $\mathscr{L}_{wz}$, touches at least one point (X_i, Y_i) for which $-1 < X_i < 0$.

Take $w < 0$ and partition the region $\{(x,y): y \leqslant x^2,\ -1 < x < 0\}$ into three parts by the two lines with respective equations $y = -(z_0 - wz_0^{-1})x + w$ and $y = (z_0 - wz_0^{-1})x + w$. Call them $\mathscr{R}_i$, for $i = 1, 2, 3$, in order from top to bottom. Let $\mathscr{E}_i$, for $i = 1, 2$, denote the event that there exists at least one Poisson point in $\mathscr{R}_i$. Conditioning on $\mathscr{E}_1$, define

$$Z_{21} = \max\{-z_0 < z < 0: \mathscr{L}_{wz} \text{ contains at least one point } (X_i, Y_i) \text{ satisfying } -1 < X_i < 0\}.$$

Writing $\mathscr{E}^{\rm c}$ for the complement of an event $\mathscr{E}$, and conditioning on $\mathscr{E}_1^{\rm c} \cap \mathscr{E}_2$, define

$$S = \min\{-(z_0 - wz_0^{-1}) < s < z_0 - wz_0^{-1}: \text{ the line with equation } y = sx + w \text{ contains at least one point } (X_i, Y_i) \text{ satisfying } -1 < X_i < 0\}.$$

Also, conditioning on $(\mathscr{E}_1 \cup \mathscr{E}_2)^c$, define

$$Z_{22} = \min\{0 < z < z_0 : \mathscr{L}_{wz} \text{ contains at least one point } (X_i, Y_i) \text{ satisfying } -1 < X_i < 0\}.$$

Each random variable Z_{21}, S and Z_{22} is well defined on the corresponding event. It may be proved as in the derivation of Theorem 2 that the probabilities $P\{W_2 \leqslant w,\ \mathscr{E}_1,\ Z_{21} \in (z, z+dz)\}$ for $-z_0 < z < 0$, and $P\{W_2 \leqslant w,\ (\mathscr{E}_1 \cup \mathscr{E}_2)^c,\ Z_{22} \in (z, z+dz)\}$ for $0 < z < z_0$, can both be expressed as $\kappa C_1^3 p_2(w,z) \exp\{-\kappa C_1^3 K_2(w,z)\}\, dz$. Also, it can be shown that the probability $P\{W_2 \leqslant w, \mathscr{E}_1^c \cap \mathscr{E}_2,\ S \in (s, s+ds)\}$ for $-(z_0 - w z_0^{-1}) < s < z_0 - w z_0^{-1}$ equals $(\kappa C_1^3/2) \exp\{-\kappa C_1^3 (2-3w)/3\}\, ds$, and does not depend on s. This completes the proof of convergence in distribution, and convergence of moments may be derived as in the case of Theorem 3.

5.3. Proof of Theorem 5

For simplicity we shall treat only the case $p = 1$.

It follows from Theorems 2 and 3 that if $h = C_1 n^{-1/3}$ then

$$n^{4/3}\, \mathrm{MSE}(x,h) = E\{W(x, C_1)^2\} + o(1), \tag{5.6}$$

where $\zeta W(x, C_1)$ denotes a random variable with distribution (in the case $x = x_0$) F_1 or F_2, according as $g''(x) < 0$ or $g''(x) > 0$ respectively. Furthermore, (5.6) holds uniformly in $C_1 \in [C^{-1}, C]$ for any $C > 1$, and in x in any compact interval $\mathscr{K}$ contained in $\mathscr{J}$. (Below we shall abbreviate the latter qualification to simply "uniformly in $x \in \mathscr{K}$".) Likewise it may be proved that

$$\lim_{C \to \infty} \liminf_{n \to \infty} \inf_{n^{1/3} h \notin [C^{-1}, C]} \inf_{x \in \mathscr{K}} n^{4/3} \mathrm{MSE}(x,h) = \infty. \tag{5.7}$$

Moreover, (5.6) and (5.7) remain true if we allow the functions g and μ to vary with n, as g_n and μ_n say, provided (a) g_n has two continuous derivatives on $\mathscr{J}$, (b) $g_n^{(j)}(x) = g^{(j)}(x) + o(1)$ uniformly in $x \in \mathscr{K}$, for $j = 0, 1, 2$, (c) $\mu_n(x,y) = 0$ whenever $y > g_n(x)$, and (d) $\mu_n^{\mathrm{cont}}(x,y) = \mu^{\mathrm{cont}}(x,y) + o(1)$ uniformly in $x \in \mathscr{K}$ and $g(x) - \varepsilon < y < g(x) + \varepsilon$, for some $\varepsilon > 0$. In (d), $\mu_n^{\mathrm{cont}}(x,y)$ denotes the "continued" version of $\mu_n(x,y)$, defined to equal $\mu_n(x,y)$ if $y < g_n(x)$ and to equal $\mu_n\{x, g_n(x)-\}$ otherwise; and μ^{cont} is defined analogously. Derivations in this very slightly more general setting are virtually identical to those in the case of fixed g and μ.

Theorem 5 follows from the convergence of moments properties discussed in Section 3 (see particularly the discussion surrounding (3.1)), (5.6), (5.7) and the following stochastic analogues of those results: for $h = C_1 n^{-1/3}$,

$$n^{4/3} \widehat{\mathrm{MSE}}(x,h) = E\{W(x, C_1)^2\} + o_p(1) \tag{5.8}$$

uniformly in $C_1 \in [C^{-1}, C]$ and in $x \in \mathscr{K}$; and

$$\lim_{C \to \infty} \liminf_{n \to \infty} P\left\{ \inf_{n^{1/3} h \notin [C^{-1}, C]} \inf_{x \in \mathscr{K}} n^{4/3} \widehat{\mathbf{MSE}}(x, h) > B \right\} = 1$$

for each $B > 0$. For brevity we shall derive only (5.8), by contradiction.

If (5.8) fails then there exists a subsequence n_ℓ, diverging to infinity, such that along no sub-subsequence of $\{n_\ell\}$ does it hold that

$$\sup_{x \in \mathscr{K}} \sup_{C_1 \in [C^{-1}, C]} |n^{4/3} \widehat{\mathbf{MSE}}(x, h) - E\{W(x, C_1)^2\}| \to 0 \tag{5.9}$$

with probability 1. (Here, $h = C_1 n^{-1/3}$.) Conditional on the data $\mathscr{P}$, the process $\mathscr{P}^*$ from which $\hat{g}^*_{\text{lin}}$ is computed is Poisson, distributed below the boundary curve $\hat{\mathscr{C}}_{\text{pil}}$. Let its conditional intensity be $\hat{\mu}_n$, and denote $\hat{g}_{\text{pil}}$ by $\hat{g}_n$ to indicate dependence on n. If we prove that

$$\text{for } j = 0, 1, 2, \quad \sup_{x \in \mathscr{K}} |\hat{g}_n^{(j)}(x) - g^{(j)}(x)| \to 0, \tag{5.10}$$

$$\sup_{x \in \mathscr{K}} |n^{-1} \hat{\mu}_n^{\text{cont}}(x, y) - \mu^{\text{cont}}(x, y)| \to 0, \tag{5.11}$$

where the mode of convergence in both cases is in probability, then it will follow that for a sufficiently sparse subsequence $n_{\ell(s)}$ of n_ℓ, we have with probability 1,

$$\text{for } j = 0, 1, 2, \quad \sup_{x \in \mathscr{K}} |\hat{g}_{n_{\ell(s)}}^{(j)}(x) - g^{(j)}(x)| \to 0, \tag{5.12}$$

$$\sup_{x \in \mathscr{K}} |n_{\ell(s)}^{-1} \hat{\mu}_{n_{\ell(s)}}^{\text{cont}}(x, y) - \mu^{\text{cont}}(x, y)| \to 0 \tag{5.13}$$

as $s \to \infty$. Let A be a subset of the sample space, satisfying $P(A) = 1$ and such that (5.12) and (5.13) hold for all $\omega \in A$. Noting the paragraph immediately below (5.7) we may deduce directly from (5.6) that (5.9) holds, in the sense of convergence for all $\omega \in A$, along the subsequence $n_{\ell(s)}$. This contradicts the claim immediately preceding (5.9), and so establishes (5.8). Therefore it suffices to derive (5.10) and (5.11).

An elaboration of the argument used to derive Theorem 2 may be employed to prove that $\hat{g}_0(y) = g(y) + O_p(n^{-2/3} \log n)$ uniformly in $y \in \mathscr{K}$. It follows from this result and the formula

$$\hat{g}_{\text{pil}}^{(j)}(x) = \frac{1}{h_1^{j+1}} \int \hat{g}_0(y) K^{(j)}\left(\frac{x - y}{h_1}\right) dy$$

that $\hat{g}_{\text{pil}}(x) = g(x) + O_p(h_1^2 + n^{-2/3} \log n)$ and, for $j = 1, 2$,

$$\hat{g}_{\text{pil}}^{(j)}(x) = g^{(j)}(x) + o(h_1^{2-j}) + O_p(h_1^{-j} n^{-2/3} \log n),$$

both results holding uniformly in $x \in \mathscr{K}$. Since $n^{(1/3)-\varepsilon} h_1 \to \infty$ for some $\varepsilon > 0$ then these results imply (5.10).

The first step in deriving (5.11) is to observe that

$$\hat{\mu}(x, y) = \sum_{Z \in \mathscr{Z}} ||\mathscr{Q}(Z)||^{-1} I\{(x, y) \in \mathscr{Q}(Z), y < \hat{g}_{\text{pil}}(x)\}, \tag{5.14}$$

where $||\mathscr{D}(Z)||$ denotes the area of $\mathscr{D}(Z)$. The conditions on h_0, and more particularly on h_1, as part of $(C_{\rm bw})$, i.e. $n^{(1-\varepsilon_1)/(p+2)}h_1 \to \infty$ and $n^{(1+\varepsilon_1)/2(p+2)}h_1 \to 0$ for some $\varepsilon_1 > 0$, ensure that for some $\varepsilon_2 > 0$ and all $B > 0$ the probability that $\sup_{x \in \mathscr{K}} |\hat{g}_{\rm pil}(x) - g(x)| \leqslant n^{-(1/2)-\varepsilon_2}$ equals $1 - O(n^{-B})$. The bound $n^{-(1/2)-\varepsilon_2}$ is of smaller order than the expected distance of a point of $\mathscr{P}$ from its nearest neighbour. These properties imply the following results.

Define $\rho(x,y) = \{n\pi\mu(x,y)/k\}^{-1/2}$. If $\varepsilon_1 > 0$ is sufficiently small then for all $B > 0$, all $\varepsilon_2 \in (0,1)$, and all $\varepsilon_3 \in (0, \frac{1}{4})$, the probability that the number of points of $\mathscr{Z}$ that lie within the disc centred on (x,y) and of radius r, is in the range $(1 \pm \varepsilon_2)n\mu(x,y)\pi r^2$ for all $x \in \mathscr{K}$, all $y \in (g(x) - \varepsilon_1, g(x) + \varepsilon_1)$, and all r satisfying $n^{\varepsilon_3 - (1/2)} \leqslant r \leqslant n^{-\varepsilon_3}$, equals $1 - O(n^{-B})$. Hence, for all $B > 0$ the probability that $d(Z)$ is in the range $(1 \pm \varepsilon_2)\,\rho(Z)$ for all $Z = (X, Y) \in \mathscr{Z}$ satisfying $X \in \mathscr{K}$ and $Y \in (g(X) - \varepsilon_1, g(X) + \varepsilon_1)$, equals $1 - O(n^{-B})$. Of course, $d(Z)$ being in the range $(1 \pm \varepsilon_2)\rho(Z)$ is equivalent to $||\mathscr{D}(Z)||$ being in the range $(1 \pm \varepsilon_2)^2 \pi \rho(Z)^2$.

Combining these results we deduce the following two properties, each holding for all $\varepsilon_1 > 0$ sufficiently small, all $B > 0$ and all $\varepsilon_2 \in (0,1)$: (a) the probability that $||\mathscr{D}(Z)||$ is in the range $(1 \pm \varepsilon_2)\{n\mu(x,y)/k\}^{-1}$ for all Z such that $(x,y) \in \mathscr{D}(Z)$ for some (x,y) satisfying $x \in \mathscr{I}$ and $g(x) - \varepsilon_1 < y < g(x) + \varepsilon_1$, equals $1 - O(n^{-B})$; and (b) the probability that the number of points $Z \in \mathscr{Z}$ such that $(x,y) \in \mathscr{D}(Z)$ is in the range $(1 \pm \varepsilon_2)n\mu(x,y)\pi\rho(x,y)^2$, equals $1 - O(n^{-B})$. Using property (a) we deduce that the right-hand side of (5.14) lies between the multiples $1 \pm \varepsilon_2$ of

$$\begin{aligned} &\{n\mu(x,y)/k\} \sum_{Z \in \mathscr{Z}} I\{(x,y) \in \mathscr{D}(Z), y < \hat{g}_{\rm pil}(x)\} \\ &\quad = \{n\mu(x,y)/k\} I\{y < \hat{g}_{\rm pil}(x)\} \sum_{Z \in \mathscr{Z}} I\{(x,y) \in \mathscr{D}(Z)\}. \end{aligned} \tag{5.15}$$

Using (b) we deduce that the series on the right-hand side of (5.15) lies between the multiples $1 \pm \varepsilon_2$ of $n\mu(x,y)\pi\rho(x,y)^2$. Each of these results holds for all $x \in \mathscr{I}$ and all $g(x) - \varepsilon_1 < y < g(x_1) + \varepsilon_1$, with probability $1 - O(n^{-B})$ for all $B > 0$, given $\varepsilon_1 > 0$ sufficiently small and any $\varepsilon_2 \in (0,1)$. Interpreting in the same sense we see, on combining the two results, that $\hat{\mu}(x,y)$ lies between the multiples $1 \pm \varepsilon_2$ of $n\mu^{\rm cont}(x,y) I\{y < \hat{g}_{\rm pil}(x)\}$. This implies (5.11).

References

[1] M. Aerts, G. Claeskens, Local polynomial estimation in multiparameter likelihood models, J. Amer. Statist. Assoc. 92 (1997) 1536–1545.
[2] J.T. Alcalá, J.A. Cristóbal, W. González-Manteiga, Goodness-of-fit test for linear models based on local polynomials, Statist. Probab. Lett. 42 (1999) 39–46.
[3] R.D. Banker, Maximum likelihood, consistency and data envelopment analysis: a statistical foundation, Management Sci. 39 (1993) 1265–1273.
[4] G. Claeskens, M. Aerts, Bootstrapping local polynomial estimators in likelihood-based models, J. Statist. Plann. Inference 86 (2000) 63–80.

[5] J.B. Copas, Local likelihood based on kernel censoring, J. Roy. Statist. Soc. Ser. B 57 (1995) 221–235.
[6] D. Deprins, L. Simar, H. Tulkens, Measuring labor inefficiency in post offices, in: M. Marchand, P. Pestieau, H. Tulkens (Eds.), The Performance of Public Enterprises: Concepts and Measurements, North-Holland, Amsterdam, 1984, pp. 243–267.
[7] J. Fan, I. Gijbels, Local Polynomial Modelling and its Applications, Chapman & Hall, London, 1996.
[8] M.J. Farrell, The measurement of productive efficiency, J. Roy. Statist. Soc. Ser. A 120 (1957) 253–281.
[9] I. Gijbels, E. Mammen, B.U. Park, L. Simar, On estimation of monotone and concave frontier functions, J. Amer. Statist. Assoc. 94 (1999) 220–228.
[10] I. Gijbels, L. Peng, Estimation of a support curve via order statistics, Extremes 3 (2000) 251–277.
[11] P. Hall, B.U. Park, S. Stern, On polynomial estimators of frontiers and boundaries, J. Multivar. Anal. 66 (1998) 71–98.
[12] W. Härdle, B.U. Park, A.B. Tsybakov, Estimation of non-sharp support boundaries, J. Multivar. Anal. 55 (1995) 205–218.
[13] N.L. Hjort, M.C. Jones, Locally parametric nonparametric density estimation, Ann. Statist. 24 (1996) 1619–1647.
[14] B.R. Jayasuriya, Testing for polynomial regression using nonparametric regression techniques, J. Amer. Statist. Assoc. 91 (1996) 1626–1631.
[15] A. Kneip, B.U. Park, L. Simar, A note on the convergence of nonparametric DEA estimators for production efficiency scores, Econometric Theory 14 (1998) 783–793.
[16] A. Korostelev, L. Simar, A.B. Tsybakov, Efficient estimation of monotone boundaries, Ann. Statist. 23 (1995) 476–489.
[17] A. Korostelev, L. Simar, A.B. Tsybakov, On estimation of monotone and convex boundaries, Pub. Inst. Stat. Univ. Paris XXXIX (1995) 3–18.
[18] C.R. Loader, Local likelihood density estimation, Ann. Statist. 24 (1996) 1602–1618.
[19] B.U. Park, W.C. Kim, J. Huh, J.W. Jeon, Estimation of density via local polynomial regression, J. Korean Statist. Soc. 27 (1998) 91–100.
[20] B.U. Park, W.C. Kim, M.C. Jones, On local likelihood density estimation, Ann. Statist. 30 (2002) 1480–1495.
[21] B.U. Park, L. Simar, C. Weiner, The FDH estimator for productivity efficiency scores: asymptotic properties, Econometric Theory 16 (2000) 855–877.